4

Certain Measures of Intergradation and Divergence

ISAAC GINSBURG

U. S. Fish and Wildlife Service,
Washington, D. C.

(Text-figure 1)

TAXONOMISTS during the pioneering stages of the science generally established and described their species on the basis of one or very few specimens. To the present day some taxonomists pursue their studies according to the methods of the pioneers. Of necessity, all of us often have to follow the same course, because of the lack of adequate material. However, the information gained by a taxonomic study of one or a few specimens of a species gives a mere smattering of our knowledge of the species. It tells us very little or nothing about variability within the species, something we need to know in order to understand properly its morphology and its affinities to the species that are immediately related to it. Therefore, by and large, the trend in modern taxonomy, and rightly so, is toward extensive studies of intraspecific or intrapopulation variability, that is, the extent and manner of individual variability within the limits of the species, or population of lower rank, with regard to those characters in which natural populations diverge.

For an adequate knowledge of the relationship and differences between closely related species of the same genus, it is desirable and necessary to know the limits and manner of individual variability of their distinguishing characters. The same holds for the minor natural populations within the species. The differences or divergences between any one pair of closely related populations, for any given character, are ascertained by determining for each individual specimen within the sample studied the numerical value of the character studied. The individual data are then tabulated, and the difference between the two populations determined by comparing the evidence presented by the two sets of tabulated data.

As studies of intrapopulation variability increase in scope and extent, it becomes evident with ever-increasing forcefulness that the gap which many taxonomists, especially those of the past, thought exists between closely related species of the same genus, is often becoming ever more narrowed or obliterated altogether. The discovery is being made repeatedly that populations which have been, and should be, afforded full species status intergrade somewhat in the very characters that have been and are being used to distinguish them. Moreover, it is being discovered that divergences between pairs of closely related populations are of all degrees of magnitude and that no sharp or definite line can be drawn between the species category and that of subspecies.

Need for These Measures.—As divergences between closely related populations are of differing extents or degrees of magnitude, it becomes imperative to have some numerical measure of the degree of divergence between a pair of populations compared for a given character. This gives the taxonomist a criterion for drawing conclusions regarding the taxonomic rank to assign to the two populations which he studies and compares, namely, whether to designate them as full species or to a particular one of the several graded intraspecific categories in the hierarchy of lesser populations. In theory, the species may be conceived as the end point in, or as comprising, a graded series of lesser populations the relationship of which increases in relative remoteness, somewhat in the following order. The most intimately knit population, consisting of a group of individuals that are the offspring of the same parents, may be designated as the family (with a small "f"). Two or more closely related families may be said to constitute a clan; two or more related clans may be said to make up a tribe. Closely related tribes theoretically constitute a variety; related varieties constitute a race; related races make a subspecies. In practice, the lesser of these minor categories are virtually altogether in the realm of theory at this stage of the science. At present we need consider only three of the upper intraspecific categories, the subspecies, the race, and the variety. In actual practice, in the great majority of instances, only the subspecies is taken cognizance of by most taxonomists. Perhaps what should more properly be designated race or variety is sometimes called subspecies. Some taxonomists use the terms race and subspecies synonymously to cover the concept of the same category.

For the present it would seem desirable to bestow a formal Latin name on the subspecies only and designate the race and variety by some informal designation, preferably one that indicates the geographic region that it inhabits, such as the "Chesapeake Bay race" or the "Potomac River variety." Perhaps in fewer instances the informal name might refer to the ecological niche that it occupies. That is, it seems best to refrain from going beyond the trinomial in nomenclature.

From the viewpoint of practicality, it is desirable that any such proposed measure have certain attributes. The taxonomist is a busy man and has to deal with thousands of populations of species rank or lower. It is, therefore, important for any such measure to be quickly determinable from the tabulated data. Also, the applicability of the derived number to the problem at hand and its pertinence in drawing acceptable taxonomic conclusions must be readily comprehensible. A proposed measure may have elegance and ingenuity from the mathematician's viewpoint, but if its derivation is too time-consuming and its pertinence not plainly evident, it will not serve as a useful tool to the taxonomist.

Some time ago I proposed two measures that seem to have the desired attributes (1938). One I called the index of intergradation, the other the index of divergence. These two indices are complementary; the greater the intergradation the less the divergence, and vice versa. For any given character compared between two populations the sum of the two indices equals 100. Some supplementary phases connected with this problem were considered by me in other papers (1937, 1939 and 1940).

Judged by private conversation with taxonomists, it seems that some misapprehensions exist in regard to the two indices as I proposed them, and the following statements seem to be in order. (1) My index of divergence cannot properly be described as the percentage of identifiable specimens in the two samples compared. (2) My index of intergradation is not represented by the percentage of the actual number of intergrading specimens in the two samples.

The paper cited above, by the use of actual taxonomic data, shows how the two proposed indices are applied in practice. The present paper discusses an underlying basis for these two indices. This may be done by considering two hypothetical populations.

Procedure.—Let us assume two natural populations, alpha and beta, that differ in a number of characters the most divergent one of which, or the principal character of which, is represented by the frequency distribution given in Table I. The class numbers might represent any character, such as the number of fin rays in a fish, the number of rays in a flower, the number of segments in the antenna of an insect, the tail length in a mouse or the length of the wing in a bird, in any chosen unit.

Table I.

Frequency Distribution of the Numerical Values of the Principal Character in Two Hypothetical Diverging Natural Populations, Alpha and Beta. (The function of the heavy vertical line, the dividing line, is discussed in the text).

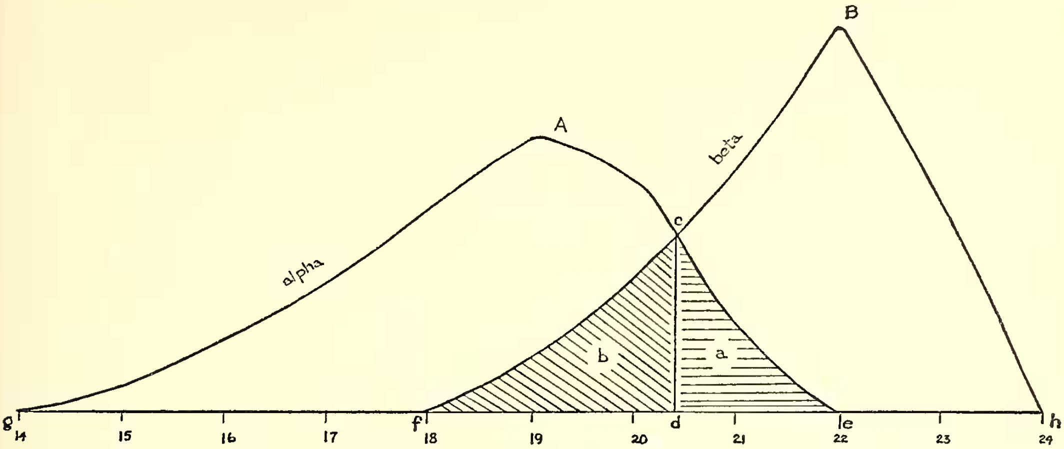
	Distribution								
Population	15	16	17	18	19	20	21	22	23
alpha	6	21	34	58	77	68	26		
beta					33	78	143	231	127

Changing the numerical frequencies in Table I to percentages of the total number of specimens in the two respective hypothetical samples, we get a distribution as given in Table II.

Table II.

Frequency Distribution of the Principal Character in Two Hypothetical Diverging Populations, Alpha and Beta, Based on the Same Assumed Data as in Table I, Except that the Frequencies Are Expressed as Percentages of the Two Respective Hypothetical Samples.

	Distribution								
Population	15	16	17	18	19	20	21	22	23
alpha	2.1	7.2	11.7	20.0	26.6	23.4	9.0		
beta					5.4	12.7	23.4	37.7	20.8



TEXT-FIG. 1.—Diagrammatic representation of the assumed data presented in Table II of two hypothetical populations, alpha and beta. Drawn by Mildred H. Carrington.

On the basis of the percentage figures in Table II, construct two overlapping curves (or polygons or histograms), as in Text-fig. 1, the curve on the left representing alpha, that on the right beta.

It should be stated at this point that the method here discussed is generally applicable to data derived from fairly homogeneous populations and based on fairly adequate samples. The comparison of two such sets of data from two closely related populations will usually result in two curves that intersect at one point. This will apply to the great majority of instances that may be encountered in taxonomic practice. In the relatively few special instances that may be expected, in which the distributions are irregular, and the curves intersect in more than one point, some modification of this method might be necessary.

Let the area enclosed by the curve and base line representing alpha be denoted by A and that representing beta by B.

From the point of intersection of the two curves draw a vertical line *cd* to the base line.

Then, the overlapped part of the graph may be considered as consisting of two—usually unequal—parts, namely, *ced* and *cfd* that may be said to represent the parts that A and B, respectively, contribute to the overlapped area.

Let areas *ced* and *cfd* be represented by *a* and *b*, respectively, *a* being longitudinally hatched in the graph and *b* obliquely hatched.

Index of Intergradation.—Now, let a′ represent the percentage of *a* into A, and b′ represent the percentage of *b* into B. Then,

$$(1)\quad a' = \frac{100\,a}{A}, \text{ and}$$

$$(2)\quad b' = \frac{100\,b}{B}$$

Combining equations (1) and (2) we have,

$$(3)\quad a' + b' = \frac{100\,a}{A} + \frac{100\,b}{B}$$

But B =A, since both curves are constructed on a percentage basis. Therefore,

$$(4)\quad a' + b' = \frac{100\,(a+b)}{A}$$

Dividing both sides of equation (4) by 2, we have

$$(5)\quad \frac{a' + b'}{2} = \frac{100\,(a+b)}{2A}$$

But $\frac{a' + b'}{2}$ equals my index of intergradation by definition. This means that this index of intergradation, when graphically presented, equals the percentage of the area overlapped by the two curves into the sum of the areas of both curves constructed on a percentage basis.

Index of Divergence.—Regarding my index of divergence, by definition it equals 100 minus the index of intergradation, that is,

$$\text{Index of divergence} = 100 - \frac{100\,(a+b)}{2A}$$

$$= \frac{100\,[2A - (a+b)]}{2A}$$

This means that, graphically presented, the index of divergence equals the percentage of the combined areas of the two curves diminished by the overlapped area, into the combined areas of the two curves. In other words, by reference to Text-fig. 1, it equals the percentage of areas *cgd* + *chd* into A + B.

Practical Use of Indices.—I have used these two indices with satisfactory results in drawing taxonomic conclusions in the following papers cited below: 1944: 376; 1950: 504 and 518; 1951: 198 (on last page cited "index of divergence" erroneously printed for "index of intergradation"); 1952: 94; 1953: 35. In addition to the pages cited where the indices are formally applied, I applied them tacitly in drawing taxonomic conclusions in other places, from the determined data. In general, I found the idea of the use of these indices and their practical application of much help in the pursuit of my studies.

In my original paper (1938), I suggested two ways of calculating the index of intergradation in practice. By one method the following steps are taken. (1) Arrange the data for the two populations compared in a frequency distribution table. (2) From the arranged data in the table construct two overlapping curves representing the two respective populations. (3) Draw a vertical dividing line in the table at a point corresponding to the point of intersection of the two curves, as the heavy vertical line in Tables I and II, the dividing line in our hypothetical example being between the classes 20 and 21. (4) Calculate the percentage of those intergrading specimens that cross the dividing line in each distribution, namely, the specimens to the right of the dividing line in alpha and those to the left of the line in beta, into the total number of specimens in each respective distribution. (5) Add the numbers thus obtained and divide by two. Another way is to tabulate the frequencies in the form of percentage as in Table II, add the smaller of the overlapping frequencies and divide the sum by 2. Applying either method to the assumed data here considered, the index of intergradation shown by the two sets of data is 14 and is of subspecies magnitude, near the lower limit of that magnitude, as previously proposed by me (1938).

The index of divergence may be calculated in two ways complementary to that in which the index of intergradation is calculated, as follows: (1) By reference to Table I, calculate the percentage of the specimens of alpha that are to the left of the dividing line, into the total number of specimens in that sample. Likewise, calculate the percentage of the specimens of beta that are to the right of the dividing line, into the total number of specimens in its sample. Add the two numbers thus obtained and divide by 2. (2) By reference to Table II, add the percentages of the frequencies to the right and to the left of the dividing line, for alpha and beta, respectively, and divide by two. Another way to determine the index of divergence is to first determine the index of intergradation and subtract it from 100. Whichever method is used, the index of divergence of alpha and beta is 86, which is of subspecies magnitude according to the schedule previously proposed by me (1938), near the higher end of that magnitude.

Two Alternative Procedures.—Instead of expressing the index of intergradation as the percentage of the overlapped area into the sum of areas A and B, it may be expressed as the percentage of the overlapped area into one of these areas, namely A, which equals B. In that case the index of intergradation is expressed by the above equation (4), and the step of dividing by 2, as in equation (5), is omitted. The effect of this procedure is to have this index doubled in its numerical value.

Starting with this latter possible index of intergradation, two possible indices of divergence may be proposed.

One method is to propose an index of divergence as represented by the following equation:

Index of divergence = 200—Index of intergradation

$$= 200 - \frac{100\,(a+b)}{A}$$

$$= \frac{100\,[2A - (a+b)]}{A}$$

This possible index of divergence then represents the percentage of the sum of the areas of the two curves diminished by the overlapped area, into one of the curves. The effect of this procedure is to double the numerical value of the index of divergence also, as compared with the procedure originally proposed and discussed above.

The second possible method is to have the index of divergence represented by the following equation:

Index of divergence = 100—Index of intergradation

$$= 100 - \frac{100\,(a+b)}{A}$$

$$= \frac{100\,[A - (a+b)]}{A}$$

This latter possible index of divergence then is represented as the percentage of the area of one of the curves diminished by the overlapped area, or in other words, as the percentage of the area of that part of one curve that is not overlapped, into the area of one curve.

Which one of the above three methods, the one originally employed and the possible two here suggested, is best, is not clear to me at present. Perhaps this could be ascertained by their application to an adequate series of comparable pairs of populations. As these indices merely have a relative, instead of an absolute, value, it is probable that it would not make much difference which one is employed as long as the same one is used consistently throughout for comparative purposes. The one originally proposed and considered at greater length above seems to be rather more compact and expressive and is perhaps preferable. In reality the three methods are merely modifications of the same basic method. They are all based on a determination, in different ways, of the relation of the overlapped area to the area of one curve or both curves combined.

It is interesting to note that by the use of the original method, the index of intergradation cannot exceed 50, while the index of divergence cannot be less than 50. By the first one of the two alternative methods the index of intergradation runs the gamut of 0 to 100; while the index of divergence cannot be less than 100. By the second of the alternative methods both indices will differ in value from 0 to 100.

Short Cut.—Expressing the frequencies in the form of percentages, as in Table II, is advantageous in that the various numbers concerned in the analysis of Text-fig. 1 may be readily determined directly from the table. Thus, the value of a equals the sum of the values in alpha that are located to the right of the dividing line which in the hypothetical example equals 9. The value of b equals the sum of the values in beta to the left of the dividing line, namely, $12.7 + 5.4 = 18.1$. The numerical value of the part of the curve that is not overlapped equals $100-(a + b) = 100-(9 + 18.1) = 72.9$.

Acknowledgment.—I wish to express my grateful appreciation to Arthur Schach and to Ralph P. Silliman who read successive trial versions of the manuscript and made constructive suggestions for its clarification.

Literature Cited

Ginsburg, Isaac

1937. The species and its subdivisions. Copeia, 1937 (3): 184-188.

1938. Arithmetical definition of the species, subspecies and race concept, with a proposal for a modified nomenclature. Zoologica, 23 (3): 253-286.

1939. The measure of population divergence and multiplicity of characters. Jour. Washington Acad. Sci., 29 (8): 317-330.

1940. Divergence and probability in taxonomy. Zoologica, 25 (1): 15-31.

1944. A description of a new gobiid fish from Venezuela, with notes on the genus Garmannia. Jour. Washington Acad. Sci., 34 (11): 375-380.

1950. Review of the western Atlantic Triglidae (Fishes). Texas Jour. Sci., 2 (4): 489-527.

1951. Western Atlantic tonguefishes with descriptions of six new species. Zoologica, 36 (3): 185-201.

1952. Eight new fishes from the Gulf coast of the United States, with two new genera and notes on geographic distribution. Jour. Washington Acad. Sci., 42 (3): 84-101.

1953. Western Atlantic scorpionfishes. Smithsonian Misc. Coll., 121 (3): 1-103.